ECONOMIC ACTIVITY FRAUDS ARE DETECTED USING MACHINE LEARNING

Prof. Mridula Shukla[1*], Sufiya Ali M[2], Srilakshmi C[3], Sowmya M S[4], Shilpa N[5] , Supriya V[6]

Article History: Received: 15.08.2022 **Revised:** 16.10.2022 **Accepted:** 22.12.2022

Abstract

The article discusses fraud detection and how to fully automate it. It is now crucial for fraud detection in every bank. There is a considerable increase in fraud, which causes the banks significant losses. Transactions present particular difficulties for fraud exposure because there is no short-term processing available. A feasibility analysis of the selected fraud detection systems is the first task. These transactions are to be tested individually and continued with the aid of models. We first establish a detection task, including the dataset's characteristics, the chosen metric, and any controls for such unbalanced datasets. This leads to the discovery that the dataset's underlying pattern produces the following results:For instance, cardholders may alter their buying patterns over time, while fraudsters may alter their strategies. Later, we presented a number of techniques for obtaining credit card sequential features. Financial fraud is the practise of acquiring financial benefits by dishonest and illegal ways. Financial fraud has recently become a severe concern to businesses and organisations, which is defined as the employment of dishonest means to earn financial profits. Financial fraud continues to have a detrimental influence on society and the economy despite numerous efforts to stop it because the daily losses from fraud amount to substantial sums of money. Many years ago, the first fraud detection techniques were established. The bulk of outdated processes still use manual labour, which is not only expensive, imprecise, and time-consuming, but also unworkable. More studies are being conducted, however they have no effect in lowering fraud-related losses. This study uses the Random Forest Classifier Machine Learning Algorithm to provide a novel model of fraud detection on bank payments. We have shown that our proposed system, which uses the Banksim dataset, is superior to the existing one by reaching train and test accuracy of 99%.

Keywords: Machine Learning, Fraud Detection, Random Forest.

[1*,2,3,4,5,6]Dept. Of MCA, the Oxford College of Engineering, Bengaluru, Karnataka, India- 560068

Email: [1*]mridula.tewari@gmail.com, [2]sufiyaalimmca2023@gmail.com, [3]srilakshmimca2023@gmail.com, [4]shilpanmca2023@gmail.com, [5]sowmyamsmca2023@gmail.com, [6]supriyavmca2023@gmail.com

[*]Corresponding Author: [1*]Prof. Mridula Shukla
Email: [1*]mridula.tewari@gmail.com
[1*]Dept. Of MCA, the Oxford College of Engineering, Bengaluru, Karnataka, India- 560068

DOI: 10.31838/Ecb/2022.11.12.53

1. Introduction

The term "deep learning" refers to a set of AI algorithms that can learn from mistakes and get better. Itself without having explicit programming. Artificial intelligence includes machine learning, which uses statistical methods combining data to forecast a result that may be used to produce useful insights. The idea behind the invention is that a computer can learn from data (i.e., examples) and provide correct results all on its own. Machine learning is closely connected to mining data and Bayesian forecasting. The computer processes info as input and uses an algorithm to produce results. "Machine learn" refers to a set of AI methods that can learn from mistakes and get better. without having explicit programming. Artificial intelligence includes machine learning, which uses statistical methods and data utilised to forecast results that may be used to generate actionable insights. The idea behind the invention is that a computer can learn from data (i.e., examples) and provide correct results all on its own. Machine learning is closely connected to mining data and Bayes forecasting. The computer processes data in input and uses an algorithm to produce results. The brain of All learning happens in AI. A machine learns in a manner similar to how someone learns. People learn by experience. Forecasting becomes simpler as our knowledge increases. By analogy, when we meet an unknown circumstance, our odds of success are fewer than they are in a regular one. The identical instruction is given to machines. In order to make an exact forecast, the computer looks at an example. When we give it a scenario that is similar, it can anticipate the outcome. But when given a fresh dataset, it struggles to forecast, just like a person. example. Learning and inference are at the heart of machine learning. The first way the machine learns is by identifying patterns. This conclusion was made possible by the facts. One of the most crucial talents of a data scientist is their ability to carefully choose the data they feed the computer. feature vector is a collection of properties used to address a problem. A feature vector is a subset of info that is used to address a challenge. Using some advanced techniques, the computer simplifies reality and turns it into a model. As a result, during the learning phase, the data are summarized and expressed as a model.

Literature Survey

Delecourt, S., and Guo, [1] L.In many nations, mobile payments are taking over as the primary payment option. Mobile payment fraud is more common than credit card fraud, though. One possible explanation is that mobile data is more easily altered by fraudsters than credit card data, which weakens our data-driven fraud detection system. Methods of supervised learning are widely applied in the investigation of fraud. These supervised learning techniques for fraud detection, however, have usually been created with the presumption that the environment is benign and that there are no adversaries attempting to defeat the system. In order to create a reliable mobile fraud detection system utilising adversarial instances, we considered various fraudster responses in this article. Experimental findings demonstrated that the effectiveness of our. M. Anwer and A. M. AlsubaieIn [2] order to detect all potential differences in the data sets and develop the statistical procedures most appropriate to remove these inconsistencies depending on the source of each discrepancy, this study presents a thorough examination of 30-minute data sets of KSA residential digital metres. Through the use of a Python-Pandas programme, the analysis is carried out. The programme analyses 3,283 users' metre readings over the course of three months in KSA and looks for data inconsistencies, duplication, missing values, outliers, and other problems. The program's statistical algorithms are then used to account for these problems. To guarantee that the adjustment process results in the most trustworthy results, a validation method was created and integrated into the programme. H. Jaiman and S. Lawte [3] With the rise of e-commerce and online transactions in the current day, credit card fraud is a critical and expanding issue. Such nefarious activities have the potential to have a huge global impact on millions of people through identity theft and financial loss. The financial system is increasingly under threat from criminal activities, which has wide-reaching effects. The efficacy of fraud detection in credit card purchases is highly impacted by the data set measuring method, the choice of variable, and the detection algorithms used. Information extraction appeared to have played a basic role in the recognition of online payment fraud. The execution of Support Vector Machine, Naive Bayes, Logistic Regression, and K-Nearest Neighbour on highly distorted data related to credit card fraud is examined in this publication. Mr. Agrawal [4] Credit card usage has grown incredibly widespread in the current economic climate. These cards make it possible for the user to make significant payments without having to carry a lot of cash. They have revolutionised cashless transactions and made it simple for customers to make payments of any kind. Although incredibly helpful, this electronic payment method carries a unique set of dangers. The number of credit card scams is rising at a similar rate to the number of users. An individual's credit card details may be unlawfully obtained and applied to fraudulent transactions. To solve this issue, Certain ways of machine learning could be employed to gather data. A. Kumar and M. Agrawal [5] Online purchases are frequently made using credit cards. There have been reports of frauds utilising credit cards in recent years. The fraud committed with a credit card is exceedingly challenging to find and stop. Many problems in science and engineering are solved using the Artificial Intelligence (AI) technique ML stands for machines learning. In this study, machine learning algorithms are used to analyse a data set of credit card frauds, and the abilities of three machine learning algorithms to identify credit card fraud are compared. In comparison to Decision Tree and XGBOOST methods, the accuracy of the Random Forest machine learning algorithm is the highest.

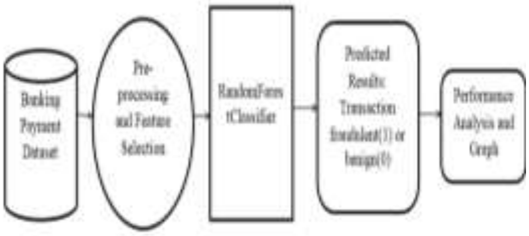

Banking Payment Dataset
Pre-processing and Feature Selection
RandomForestClassifier
Predicted Results: Transaction fraudulent(1) or benign(0)
Performance Analysis and Graph

Fig. 1. Proposed Architecture

Existing Model

- Using statistical methods, Abdallah et al. introduced a review to look into several strategies for spotting fraud in the healthcare industry.
- Popat and Chaudhary offered a thorough study of the field of detecting credit card fraud. The authors offer a thorough examination of numerous ML classification algorithms, including their methodology and difficulties.
- Ryman-Tubb et al. investigated a number of cutting-edge techniques for identifying payment card fraud using transactional volumes. Only eight ideas have a realistic application for use in the sector, according to the report.
- A research by Albashrawi and Lowell examined numerous studies over ten years that used data mining techniques to detect fraud in the banking sector.
- The system model that is currently in use with logistic regression is unable to forecast a continuous result.
- If the sample size is too little, the current system model with logistic regression may not be reliable.
- The current system may result in an over-fitting issue.
- The quality of the data determines how accurate the current system is, and large data may slow down the prediction step.

2. Proposed Methodology

It has become exceedingly challenging for banks to identify bank payment fraud. For transactions to be free of financial fraud, machine learning is essential. In the suggested system, we employ machine learning methods to forecast these transactions. Previous data has been gathered, and new features have been included to increase the predictive power. The choice of variables, sampling strategy for the data set, and fraud detection methods all have a significant impact on how well fraud is detected in banking transactions. Bank transaction data is gathered from Kaggle.

To choose the model with the greatest accuracy, the accuracies are weighed and compared. Our suggested system model achieved test and train accuracy of 99%.

Implementation

Dataset: 594643 distinct pieces of data make up the dataset. The dataset contains 11 columns, each of which is detailed below.

The feature corresponds to the step id. Customer: The zip code of origin/source is represented by this characteristic, zipCodeOrigin.

zipThe zip merchant is 28007.

Identifier for the merchant: zipMerchant: Zip code for the merchant

Age: Age classifications include 0: = 18, 1: 19–25, 2, 36–45, 4, 46–55, 5, 56–65, and 6: > 65. The unknown

Gender: Customer's gender Enterprise, Female, Male, Unknown, and F: Female

Category: The purchase's category. Since we'll see them later in the study, I won't include all the categories here.

Quantity: The cost of the purchase

Fraud: The target variable that indicates whether a transaction is fraudulent (1) or not.

Data Collection: We create the initial module's collect data process. The real task of creating a machine learning system and accumulating data starts now. This stage is critical since the amount and quality of data we can gather will determine how effectively the model works.

Several techniques may be used to get the data, like web crawling, human interventions, and datasets stored in model folders.The dataset is taken from the well-known Kaggle dataset source. The dataset's link is provided below.

the following URL for the dataset: https://www.kaggle.com/datasets/jayaprakashpondy/banksim-dataset

Data Preprocessing: Amass data and prepare it for training. Remove duplicates, correct mistakes, manage missing values, normalise, alter data types, and anything else that may require cleaning up. Randomising the data removes the effects of the precise sequence in which we gathered and/or created our data.

Conduct further exploratory research, such as data visualisation, to find important relationships between data or class disparities (bias alert!) sets for instruction and assessment are separated.

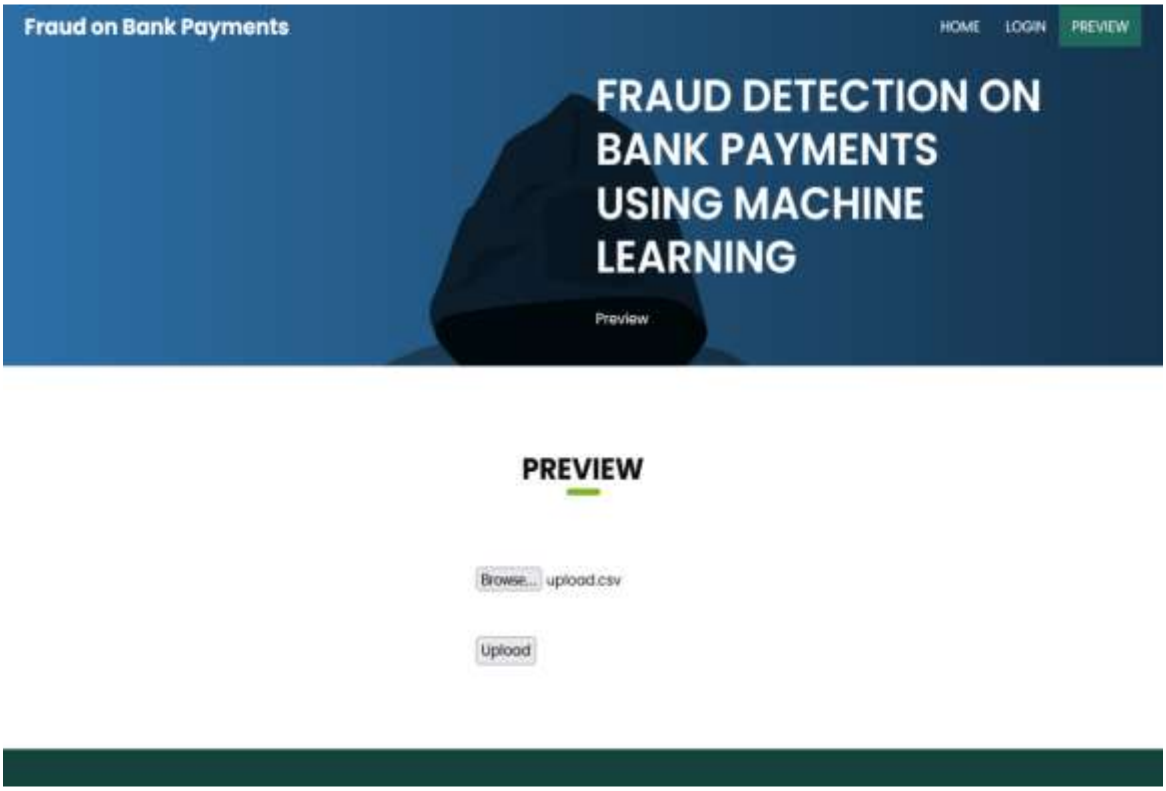
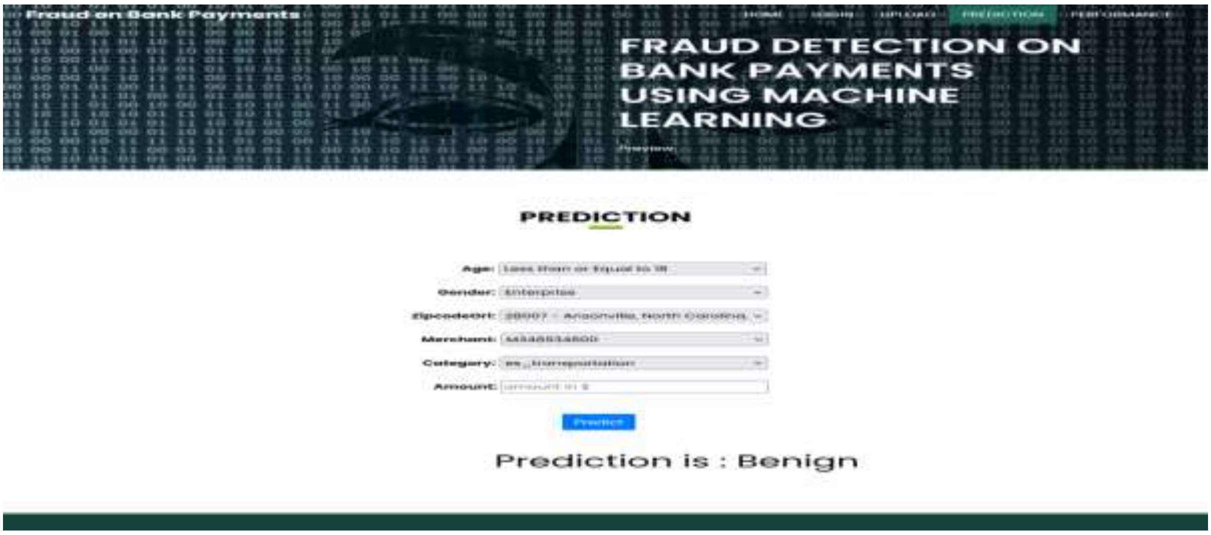


Fig 2. Preview page

Model choice

We employed The process for learning machines RandomForestClassifier. We deployed this technique after obtaining a 99.7% accuracy on the training set. Algorithm for Random Forests Let's explain the algorithm using everyday language. Let's say you want to take a trip and you want to go somewhere you will enjoy.So how can you locate a location that you will enjoy? You can perform an online search, read reviews on travel blogs and websites, or ask your friends for recommendations.

Imagine that you chose to speak with your friends in order to inquire about their earlier trips. You will receive advice from each buddy. You must now create a list of those recommended places.



PREDICTION

Age: Less than or Equal to 18

Gender: Enterprise

ZipcodeOri: 28007 - Ansonville, North Carolina

Merchant: M348934600

Category: es_transportation

Amount: amount in $

Predict

Prediction is : Benign

Fig 3. Prediction Page

Maintain the Trained Model: Once you are comfortable utilising your tested and trained model, the first task is to store it in an.h5 or.pkl file via a library like pickle a production-ready environment. Verify that Pickle is set up in your environment. It will now be saved as an a.pkl file and loaded into this module.

3. Conclusions

Various financial environments, such as the corporate, banking, insurance, and tax sectors, are susceptible to financial fraud. Businesses and industries are now concerned about financial fraud. Despite several efforts to eradicate it, financial fraud persists, which has a detrimental effect on society and the economy because everyday losses from fraud amount to extremely large quantities of money. Thanks to the advancement of artificial intelligence, machine learning-based technologies may now be utilised intelligently to detect fraudulent transactions by analysing a sizable amount of financial data.

We summarised and carefully analysed the body of research on machine learning-based fraud detection in this article. The Random Forest Classifier methodology, which is used in this study, uses well-defined methodologies to extract, synthesise, and publish results.

4. References

1. S. Delecourt and L. Guo, "Building a strong mobile payments frauddetection system with adversarial examples," 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 103-106, IEEE, 2019.
2. T. Alquthami, A. M. Alsubaie, and M. Anwer, "The Value of smartmeters for information processing - case of Saudi Arabia," in 2019 IEEE International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–5.
3. "Comparative evaluationof credit card fraud detection using machine learning techniques," in 2019 Global Conference for Advancement in Technology (GCAT), pp. 1-6, IEEE, by O. Adepoju, J. Wosowei, S. Lawte, and H. Jaiman.
4. S. Khatri, A. Arora, and A. P. Agrawal, "Supervised machine learningalgorithms for credit card fraud detection: A comparison," in 2020 10th International Conference on Cloud Computing, Data Science Engineering(Confluence), pp. 680-683, IEEE, 2020.
5. V. Jain, M. Agrawal, and A. Kumar, "Performance analysis of machinelearning algorithms in credit cards fraud detection," in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization(Trends and Future Directions) (ICRITO), pp. 86-88, IEEE.
6. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga, and N. Kuruwitaarachchi, "Real-time detection of credit card fraud using machinelearning," in 2019's 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 488-493, IEEE, 2019.